

La notion de cadre conceptuel de modélisation: penser la place des ontologies dans le processus de modélisation

Jean-Pierre Müller
CIRAD, UPR GREEN
Campus International de Baillarguet, 34398 Montpellier cedex 5, France
Jean-pierre.muller@cirad.fr

1. Introduction

L'objectif de cette contribution est de proposer la notion de cadre conceptuel de modélisation (CCM) pour revisiter le processus de modélisation informatique et y situer la place, souvent négligée des ontologies.

Un certain nombre d'auteurs se sont intéressés à la méthodologie de conception de modèle informatique et leur simulation. T. Meurisse (2004) propose d'y faire intervenir trois rôles distincts (qui peuvent ou non être joué par la même personne) : le thématicien, le modélisateur et l'informaticien. Ces distinctions lui permettent de faire intervenir trois modèles ainsi qu'il est illustré dans la figure 1.

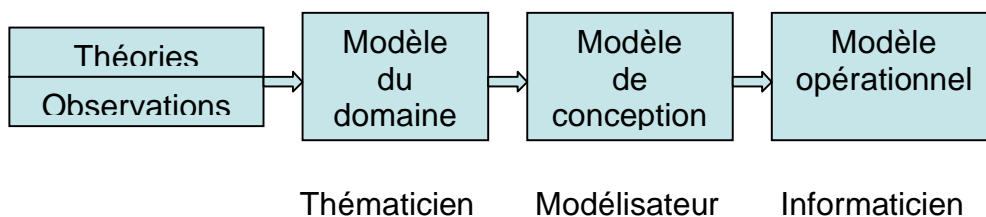


Figure 1. Méthodologie de modélisation selon T. Meurisse (2004)

Très simplement, le modèle opérationnel est le programme informatique (par exemple, un programme d'intégration numérique), le modèle de conception est exprimé dans un formalisme indépendant de l'implémentation (par exemple, des équations différentielles) et le modèle du domaine est la compréhension du thématicien (par exemple, les dynamiques proies/prédateurs). Cette proposition, très éclairante par ailleurs, possède quelques limites. D'une part, elle suggère un processus linéaire alors que les interactions sont fréquentes et fécondes entre ces différentes formulations. D'autre part, ni les liens entre théorie, observations et modèle du domaine, ni leurs outils d'expression ne sont explicités.

Selon le cadre épistémologique « sémantique » (Van Fraassen 2008), une théorie est reliée aux données empiriques à travers des modèles qui spécifient les paramètres de la théorie et l'applique à un domaine particulier : théorie – modèles – données empiriques. A chaque pas, nous avons une implication

ontologique, c'est-à-dire un présupposé sur les objets dont on parle. Qui plus, est, les discours peuvent être complètement différents, par exemple lorsqu'on passe du modèle du domaine au modèle informatique, d'où le besoin d'un médiateur entre le discours du thématicien et celui de l'informaticien. La plupart du temps, l'ontologie n'est pas formalisée et sa relation à la théorie, au modèle et aux données empiriques ne l'est pas davantage. Un certain nombre de situations typiques peuvent apparaître. Différentes théories et modèles peuvent être en compétition pour expliquer un phénomène du « monde réel » et la question est alors de leur intérêt relatif. Deux modélisateurs peuvent manipuler ce qu'ils pensent a priori comme le même objet d'investigation, alors qu'une analyse plus profonde montre des sémantiques sous-jacentes différentes. A l'opposé, différentes disciplines peuvent considérer des objets et des relations a priori différents qui peuvent, en fait, être articulés et même fusionnés. De plus, dans le même domaine académique, des programmes informatiques différents peuvent être cohérents mais sont néanmoins en compétition pour traduire un même modèle. Tous ces cas sont classiques quand on développe des modèles en sciences sociales, mais souvent, ils ne sont pas explicitement identifiés, ce qui peut donner des confusions dans l'interprétation. Dans tous les cas, rendre explicite l'ontologie sous-jacente sera une aide, ainsi que de révéler la compatibilité de modèles exprimés dans différents langages (par exemple, informatique et thématique), ainsi que de mettre en exergue la similarité ou la dissimilarité entre des modèles et des objets définis différemment (par exemple en économie et en géographie).

Afin de fournir un cadre utilisable, systématique et cohérent pour la conception de modèle, nous introduisons ce que l'on appellera un cadre conceptuel de modélisation (CCM). Il est composé de trois domaines reliés (voir la figure 2) (1) le domaine empirique, (2) le modèle du domaine, (3) le domaine conceptuel (à ne pas confondre avec le CCM qui désigne l'ensemble). Le domaine empirique est le domaine des observations et des mesures. Il dépend des pratiques d'observation (expérimentation, récolte des données, entrevues, etc.) et de la vision sous-jacente du thématicien définissant à propos de quoi sont ces observations et mesures. Nous explicitons ainsi la partie « observation » de la figure 1. Le domaine du modèle formalise les représentations (équations différentielles, assertions logiques, programmes informatiques, etc.) qui sont (comme toutes les représentations ainsi que comme les observations) à propos de ce que le thématicien veut expliquer. Ce domaine comprend donc le modèle de conception de la figure 1. Pour rendre explicite la vision sous-jacente définissant (au moins localement) à propos de quoi sont les modèles et les observations, on introduit la notion de domaine conceptuel qui est essentiellement le référentiel théorique qui sous-tend le processus de modélisation. Ce domaine conceptuel englobe à la fois la partie « théorie » et le

modèle du domaine de la figure 2. Nous laisserons momentanément le modèle informatique pour le reprendre en conclusion.

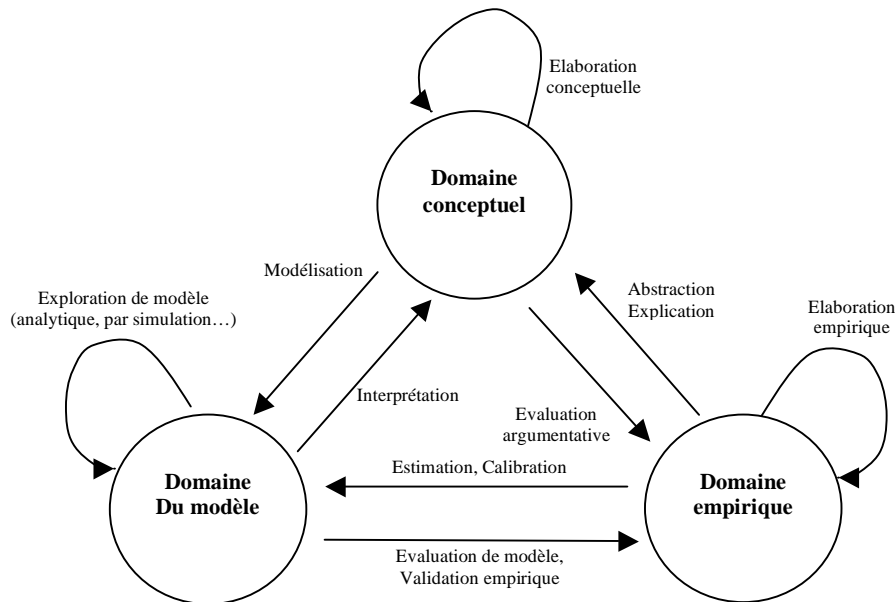
Un certain nombre de pratiques sont organisées autour de ces trois domaines. Par exemple, de nouvelles données empiriques peuvent être élaborées à partir des données primaires enrichissant ainsi le domaine empirique. Le domaine conceptuel peut être sujet à généralisation ou à spécialisation, enrichissant ainsi le domaine conceptuel. Et les modèles peuvent être explorés analytiquement ou par simulation (comme c'est souvent le cas pour les systèmes complexes), enrichissant ainsi le domaine du modèle. Les modèles peuvent être calibrés en utilisant les données empiriques et interprétés en termes des explications conceptuelles qu'ils supportent. En pratique, certains chercheurs utilisent les modèles comme un outil technique d'investigation avec seulement quelques contraintes sur les opérations formulées empiriquement, et ceci sans réelle théorie (Livet 2007) et/ou dans un cadre de modèles hétérogènes interagissant (Varenne 2007), donc, essentiellement, selon l'axe horizontal. Le domaine empirique peut faire directement l'objet d'une théorisation sans forcément utiliser des modèles, comme c'est le cas dans certains domaines des sciences humaines et sociales, et également des savoirs non-experts (ou profanes).

Dans tous les cas, il n'y a pas de raison de supposer que les ontologies sous-jacentes au domaine empirique (ce que nous observons ou mesurons), au domaine conceptuel (ce dont parle la théorie) et au domaine du modèle (ce que le modèle raconte) coïncident, ni ne soient cohérents les uns avec les autres. Il est donc important de rendre explicite les ontologies dans chaque domaine. Finalement, ces domaines peuvent être plus ou moins invoqués ou formalisés en fonction de la discipline scientifique. Ainsi, les sciences humaines privilégient, en général, plutôt le domaine conceptuel, alors que les sciences de la nature privilégient davantage le domaine du modèle.

Finalement, un processus de modélisation est toujours dépendant à la fois de la personne qui s'y attèle et de la question qu'elle se pose. Ainsi la personne (ou le groupe de personnes) qui va remplir les rôles d'observateur, de thématicien et de modélisateur ainsi que sa ou ses questions se trouve implicitement au centre de ce triangle. Réciproquement, un tel triangle va représenter un moment de son processus de modélisation qui va opéré l'un après l'autre ou conjointement sur chacune des trois dimensions. Ainsi sur un même objet d'investigation peut potentiellement s'articuler un ensemble de ces triangles qui constituent autant de points de vue qui correspondent à une multiplicité des regards portés successivement (dimension diachronique) ou conjointement (dimension synchronique).

Dans la suite nous allons élaborer davantage ce cadre conceptuel avant de donner un certain nombre de perspectives pour aider à penser le processus de modélisation.

Figure 2. Un cadre conceptuel



2. Le cadre conceptuel de modélisation

Dans cette section nous décrire en détail le processus de modélisation en utilisant le CCM qui a été informellement présenté dans l'introduction. Il reste maintenant à produire une formalisation de celui-ci afin de rendre explicite la relation entre les différents domaines. Commençons par la relation avec le domaine du modèle.

2.1. Le domaine du modèle

Sur la question “qu’est-ce qu’un modèle?”, la définition la plus citée est celle de Minsky (1965): “Pour un observateur B, un objet A^* est un modèle d’un objet A dans la mesure où B peut utiliser A^* pour répondre aux questions qu’il se pose sur A”. Un chercheur possède donc un domaine empirique d’intérêt A, appelé ci-dessus l’objet d’investigation et une question B concernant ce domaine. Pour répondre à la question B, la modélisation inclut un processus d’abstraction du domaine empirique A vers un modèle construit artificiellement A^* , avec l’idée que les entités abstraites et les relations qui structurent le modèle A^* sont “suffisantes” pour répondre à la question B sur A. La discussion qui suit se réfère à cette définition de modèle.

Un modèle est d'abord un système formel. Un système formel (SF) est défini par :

- Un ensemble V de signes (mots, lettres) et une grammaire décrivant l'ensemble, en général infini, de toutes les structures autorisées à partir de ces signes (une telle structure de signe est souvent appelée une phrase). Cet ensemble est en général appelé le langage L dans le cas de structures linéaires.
- Un ensemble de règles R pour transformer les structures entre elles. Etant donné un sous-ensemble de L , ces règles sont utilisées pour engendrer d'autres phrases du langage à partir de ce sous-ensemble. En logique, ce sous-ensemble est appelé l'ensemble des axiomes et les phrases engendrées, l'ensemble des théorèmes, mais une structure peut être aussi interprétée comme un état initial et les structures successivement engendrées comme une trajectoire dans l'espace L .

La propriété principale d'un système formel est qu'il n'a absolument aucune signification intrinsèque à moins de fournir un mécanisme interprétatif. Habituellement, une interprétation est donnée par:

- Un domaine de discours qui est relié au domaine conceptuel du thématicien ;
- Une fonction qui fait correspondre aux signes et aux phrases du langage, les éléments du domaine de discours.

De plus, les règles peuvent être interprétées comme les implémentations des lois associées à la théorie (voir plus loin). En conséquence le domaine du modèle doit se référer au domaine conceptuel pour obtenir sa sémantique. Dans la réalité, les signes ou mots ou lettres ne sont pas neutres et induisent par eux-mêmes une interprétation intuitive. C'est particulièrement le cas si les éléments du systèmes formels sont appelés des agents ou des objets comme c'est souvent le cas dans les modèles informatiques. Si nous voulons prendre en compte les biais interprétatifs entre ce que le modèle peut signifier et la théorie du thématicien, il n'est donc pas inutile d'inclure l'interprétation du modèle dans le domaine du modèle lui-même. Le domaine de discours n'étant rien d'autre qu'une ontologie, nous allons donc en formaliser l'ontologie.

Formellement, une ontologie dans le sens philosophique est une paire $\langle E, P \rangle$ où E est un ensemble d'entités catégorisées, appelés aussi individus, et P est un ensemble de propriétés nommées (ou qualités) et de relations sur E . En informatique, une ontologie est habituellement divisée dans un modèle conceptuel fait de concepts ou catégories structurées par des relations taxonomiques et sémantiques et d'un modèle concret fait d'objets qualifiés ou individus structurés par des liens. La même distinction peut être trouvée dans les logiques descriptives (Baader 2003) où les formules sont divisées en une boîte terminologique (ou T-Box) correspondant au modèle conceptuel et une boîte d'assertions (ou A-Box) qui correspond au modèle concret. La première boîte

définit le vocabulaire utilisé pour décrire la seconde. La paire $\langle E, P \rangle$ correspond au modèle concret (ou A-Box) mais nous utiliserons dans la suite la définition étendue d'ontologie telle qu'utilisée en informatique. Finalement l'interprétation d'un modèle n'est rien d'autre qu'une fonction faisant correspondre aux structures L du système formel des sous-ensembles du modèle conceptuel (ou A-Box), et, plus finement, à chaque signe dans V , un élément de E . En conséquence, un système formel peut être interprété de différentes façons selon la relation que l'on construit avec l'ontologie. De la même façon, une même ontologie peut donner sens à différents systèmes formels.

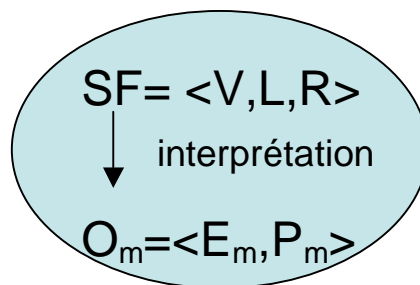


Figure 3. Le domaine du modèle

2.2. Le domaine conceptuel

Le domaine conceptuel peut lui-même être formalisé sous la forme d'une ontologie et d'une théorie. Une théorie T est un ensemble de contraintes et déterminations (relations théoriques ou lois) sur l'ontologie, sous la forme d'une paire $\langle C, G \rangle$ où C est un ensemble de comportements, et possiblement de déterminations causales et G est l'ensemble des déterminations et contraintes descriptives générales (par exemple, la composition et/ou la conservation, etc.) exprimées sur les propriétés des entités concrètes. Nous appellerons donc la combinaison d'une ontologie et d'une théorie un domaine conceptuel.

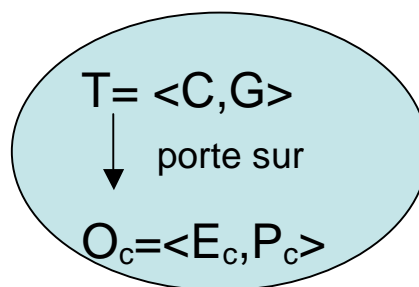


Figure 4. Le domaine conceptuel

Nous pouvons maintenant construire la relation entre le domaine conceptuel et le domaine du modèle. Le domaine de discours sur lequel le système formel acquière sa sémantique est formalisé par le modèle concret (ou A-Box) de

l'ontologie et donc par O_m . Si ce dont parle l'informaticien a une quelconque correspondance avec le discours du thématicien, il doit exister une relation entre O_c et O_m . Dans le meilleur des cas, cette relation est l'identité pure et simple et donc, il existe un sens strict pour lequel le modèle est un bel et bien le modèle du thématicien. En pratique, une relation plus lâche est possible par le seul fait qu'un modèle n'est pas nécessairement assez riche pour exprimer tout ce que le thématicien pourrait exprimer. Cela donne, au passage, un sens formel à l'expressivité d'un système formel. De la même façon, l'ensemble R des règles du système formel peut être mise en correspondance avec tout ou partie de la théorie T , permettant ainsi au système formel de rendre compte des raisonnements que le thématicien opère au sein de son modèle conceptuel. Cette correspondance est formalisée en logique par les notions de complétude et de consistance : c'est-à-dire la propriété d'un système logique de produire tout (complétude) et que (consistance) ce que le thématicien tient pour vrai étant donné sa théorie.

2.3 Le domaine empirique

Finalement, le domaine empirique est l'élaboration des données proprement dites, qu'elles soient des mesures (quantitatives ou qualitatives), des énoncés ou n'importe quelle forme. Ces données peuvent être obtenues par des méthodes diverses : observations directes, expérimentation, interviews, recherches bibliographiques (faisant ainsi la distinction entre données primaires et secondaires). Ce qui caractérise des données, c'est leur interprétation comme mesures ou observations de quelque chose, c'est-à-dire d'une « réalité » fait d'objet et de phénomènes dont ces données constituent les traces. De nouveau, l'observateur ne peut se défaire d'un engagement ontologique. Aussi, proposons nous de formaliser le domaine empirique par la définition des données D d'une part, et de l'ontologie de l'observateur O_o . Nous introduisons une troisième ontologie, car, encore une fois, il n'est pas nécessaire de supposer que l'ontologie de l'observateur coïncide avec les ontologies du thématicien, respectivement du modélisateur, même si, à terme, nous souhaitons leur convergence.

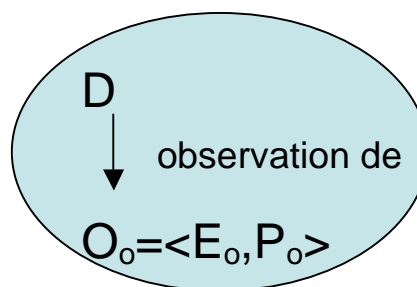


Figure 5. Le domaine empirique

En mettant en correspondance l'ontologie d'observation avec respectivement les ontologies conceptuelles et du modèle, il est possible alors d'affiner les aller-retour entre l'empirique, le théorique et le formel, affinant ainsi le processus de modélisation, voire de construction des connaissances qu'elles soient disciplinaires ou profanes.

2.4. Synthèse

L'introduction d'un CCM systématique constitue une systématisation formelle du rôle de l'ontologie dans le processus de construction d'un modèle. La figure 2 synthétise cette approche. L'objectif est d'introduire une approche systématique sur le processus de modélisation associant le thématique, l'observateur et le modélisateur. C'est un processus à trois pôles, impliquant le domaine empirique auquel on s'intéresse, le domaine conceptuel, et le domaine du modèle. Un CCM particulier est donc une façon spécifique de systématiser la relation entre les trois pôles en utilisant les ontologies et la théorie comme des outils centraux pour spécifier une conceptualisation particulière du domaine. Dans ce processus de modélisation, aux observations du thématique sur le domaine empirique (par exemple, les données empiriques) doit être attribué une signification par un processus de thématisation. Ce mécanisme d'abstraction attribue ces observations à des objets, des relations et des propriétés qui doivent être défini dans le domaine conceptuel (ontologie). Ces éléments ontologiques sont aussi sujet à des relations nomologiques qui introduisent des contraintes sur les possible opérations dans le modèle conceptuel résultant (ou « théorie » dans le sens large utilisé dans l'introduction). Par facilité, ces relations ont été appelées « lois » correspondant au domaine conceptuel (ou « théorie »). La combinaison à la fois d'une ontologie et d'une théorie exprime le domaine conceptuel particulier sur le domaine empirique que le modélisateur – où l'analyste – extrait des observations du thématique. Il est résumé sous la forme abstraite d'un ensemble d'entités, propriétés et relations constitutives de cette ontologie et de l'ensemble des lois constitutives de cette théorie. Cette vue conceptuelle doit ensuite être implémenté dans le domaine des modèles. C'est la tâche du processus de conception. Par ce processus, les relations théoriques (« lois ») sont implémentées sous la forme de règles d'un système formel et le système formel et le domaine conceptuel partage une même ontologie de base dans laquelle le système formel puise sa sémantique.

Dans la figure 6, nous avons fait l'hypothèse que chaque pratique (de l'observateur, du thématicien et du modélisateur) induisait une ontologie distincte. Mais des convergences partielles sont possibles et surtout nécessaires, et d'abord entre le modèle conceptuel et celui du modèle. Comme il est mentionné dans l'introduction, un schéma plus général a été considéré avec une ontologie différente dans chaque domaine de façon à rendre explicite le processus de convergence à réaliser entre l'observation, la théorie et le modèle. Une autre possibilité, souvent utilisée dans les sciences de la Nature, est de construire directement les modèles dont le mécanisme est ajusté pour correspondre aux observations (calibration) et pour prédire d'autres observations (validation). Dans ce cas, à la fois les règles (R) et l'ontologie (O) doivent nécessairement représenter le processus empirique, entités et relations, mais seulement pour être efficace dans les prédictions, ainsi qu'il est argumenté par

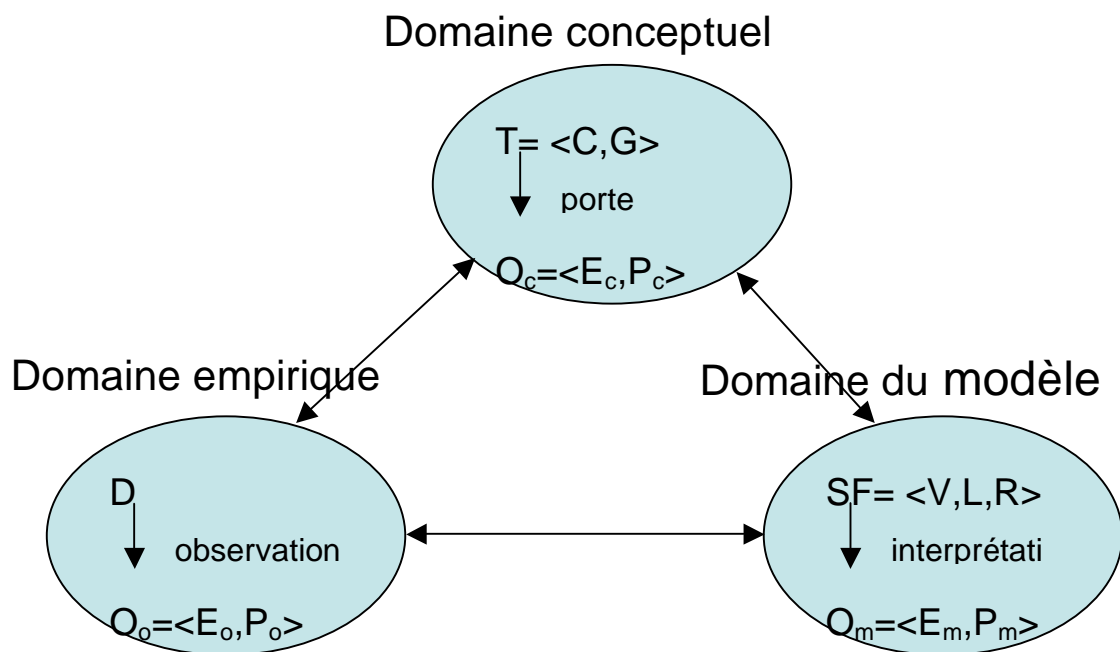


Figure 6. L'ontologie dans le processus de conception du modèle au sein d'un CCM

les instrumentalistes. Cependant, dans d'autres cas, le modèle lui-même peut être utilisé pour construire a posteriori un domaine conceptuel. Dans ce cas, les ontologies correspondantes sont très proches du contenu du modèle. Inversement, dans cette approche par le cadre de connaissance, les relations nomologiques et l'ontologie sont également utilisées comme un point de départ pour concevoir les modèles. Si un modèle est intrinsèquement une représentation formelle, ce n'est pas nécessairement le cas pour les ontologies et les théories, d'où l'intérêt pour les ontologies informatiques comme représentations pour exprimer les ontologies thématiques qui sont quasi absentes

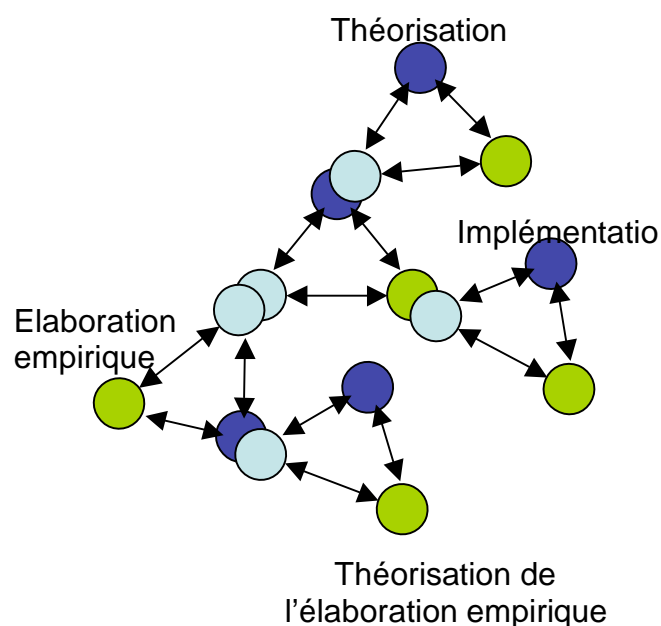
dans les sciences de la Nature et qui sont, à peu près, les seuls objets du discours dans les sciences humaines et sociales.

Conclusion et perspectives

Nous avons donc proposé l'esquisse d'une approche formelle du processus de modélisation prenant en compte à la fois le domaine empirique, le domaine conceptuel ou théorisant et le domaine du modèle proprement dit sous la forme d'un cadre conceptuel de modélisation (CCM). Ce n'est qu'une esquisse puisqu'il nous faudrait maintenant préciser les opérations permettant de passer d'un domaine à l'autre et de définir, ce que veut dire des ontologies cohérentes, contradictoires, etc. médié par des opérations de mise en correspondance des vocabulaires définis par les modèles conceptuels des ontologies.

Il reste également à élargir ce cadre jusqu'à l'implémentation de façon à couvrir les discours méthodologiques tels que celui de T. Meurisse (2004). Pour ce faire, nous pensons qu'une piste possible est de considérer chacun des trois domaines comme un domaine empirique modélisable. Ainsi, on pourrait prendre le système formel du domaine du modèle et en faire une théorie et un modèle dans un nouveau système formel qui ne serait rien d'autre qu'un programme informatique qui essaye de réaliser le système formel initial (et peut faire lui-même l'objet d'une théorisation). On peut faire de même avec les domaines empirique et conceptuel afin de modéliser l'élaboration des données ainsi que l'élaboration théorique. Une telle généralisation est illustrée par la figure 7.

Figure 7. Généralisation du CCM



Bibliographie

- Baader F. (2003), “Description Logic Terminology”, dans: Description Logic Handbook Theory, Implementation, and Applications Cambridge University Press, 2003, p.485-495.
- Livet Pierre “Towards an Epistemology of Multi-agent Simulation in Social Science”, in Phan, Amblard. (2007) op.cit. p. 169-194.
- Meurisse Thomas (2004), Simulation multi-agent : du modèle à l’opérationnalisation, Thèse de l’Université Paris VI, juillet 2004, LIP6
- Van Fraassen Bas C. (2008), Scientific Representation : Paradoxes of Perspective, Oxford University Press, 2008
- Varenne Frank (2007), Du modèle à la simulation informatique, Paris, Vrin, 2007